UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life

Huynh, Elisabeth; Coast, Joanna; Rose, John; Kinghorn, Philip; Flynn, Terry

DOI:
10.1016/j.socscimed.2017.07.012

License:
Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Huynh, E, Coast, J, Rose, J, Kinghorn, P & Flynn, T 2017, 'Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life', *Social Science and Medicine*, vol. 189, pp. 114-128. https://doi.org/10.1016/j.socscimed.2017.07.012

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 04. May. 2023

Accepted Manuscript

Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life

Elisabeth Huynh, Joanna Coast, John Rose, Philip Kinghorn, Terry Flynn

PII: S0277-9536(17)30445-8

DOI: 10.1016/j.socscimed.2017.07.012

Reference: SSM 11322

To appear in: *Social Science & Medicine*

Received Date: 19 December 2016

Revised Date: 15 May 2017

Accepted Date: 19 July 2017

Please cite this article as: Huynh, E., Coast, J., Rose, J., Kinghorn, P., Flynn, T., Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life, *Social Science & Medicine* (2017), doi: 10.1016/j.socscimed.2017.07.012.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Original article

VALUES FOR THE ICECAP-SUPPORTIVE CARE MEASURE (ICECAP-SCM) FOR USE IN ECONOMIC EVALUATION AT END OF LIFE

Running title: Values for ICECAP-SCM

Elisabeth Huynh

Institute for Choice, University of South Australia, Australia

Joanna Coast *

School of Social & Community Medicine, University of Bristol, UK

John Rose

University of Technology Sydney, Australia

Philip Kinghorn

Health Economics Unit, Institute of Applied Health Research, University of Birmingham, UK

Terry Flynn

TF Choices Ltd, Nottingham, UK

**** Please address correspondence to:***
Joanna Coast,
Professor in the Economics of Health & Care
School of Social & Community Medicine,
University of Bristol
Canynge Hall, 39 Whatley Road
Bristol BS8 2PS
Tel: 0117 928 7387
E-mail: jo.coast@bristol.ac.uk

KEYWORDS: UK; capability; measurement; ICECAP measures; wellbeing; end of life; best-worst scaling; discrete choice experiment.

FUNDING STATEMENT: An EU funded ERC Starting Grant (261098 EconEndLife) funded Joanna Coast and this work to conduct research into the allocation of resources in end of life care using a capability framework.

CONFLICT OF INTERESTS DECLARATION: JC led the development of the ICECAP-SCM.

ACCEPTED MANUSCRIPT

HIGHLIGHTS

- First paper to provide tariff values for ICECAP Supportive Care Measure
- Two sets of ICECAP-SCM values generated, one of which includes interaction terms
- Feasible to use these new values in economic evaluations of interventions at the end of life

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

SUMMARY

End of life care may have elements of value that go beyond health. A generic measure of the benefits of end of life care could be helpful to decision makers. Such a measure, based on the capability approach, has recently been developed: the ICECAP Supportive Care Measure. This paper reports the first valuation exercise for that measure, with data from 6,020 individuals collected from an on-line general population panel during June 2013. Individuals were asked to complete a stated choice experiment that combined best-worst scaling and a standard discrete choice experiment. Analysis of the best-worst data used limited dependent variable models within the random utility framework including the multinomial logit models and latent class choice model analysis. Exploratory steps were taken to determine the similarity of the best-worst and DCE data before formal testing and pooling of the two data sources. Combined data were analysed in a heteroscedastic conditional logit model adjusting for continuous scale. Two sets of tariffs were generated, one from the best-worst data capturing only main effects, and a second from the pooled data allowing for two-way interactions. Either tariff could be used in economic evaluation of interventions at the end of life, although there are advantages and disadvantages with each. This extensive valuation exercise for the ICECAP Supportive Care Measure with a large number of members of the general public could be complemented in the future with best-worst scaling studies amongst those experiencing the end of life.

KEYWORDS: UK; capability; measurement; ICECAP measures; wellbeing; end of life; best-worst scaling; discrete choice experiment.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

INTRODUCTION

In the economics of health and care, care at the end of life is increasingly recognised as involving elements of value that go beyond gain in health. Across academic health economics (Normand, 2009, 2012) (anonymous1), popular medical writing (Gawande, 2014, 2016), and national policy (Canada, 2007; Department of Health, 2008; Department of Health and Ageing, 2010; Parliamentary Assembly, 2009), there is increasing realisation that focusing just on gains in health at this point in the life course, usually considered to be the last year of life (although this, of course, may be difficult to predict), may be much too narrow and restrictive (anonymous1, anonymous5). Equally limiting, however, is an over-emphasis in policy and much practical research in end of life care, on place of death as the sole outcome of end of life provision. Place of death is often used (implicitly or explicitly) as a proxy for a good (home) or bad (hospital) death, but this too is an oversimplification. More valuable, would be a measure of the benefits of end of life care, that is not specific to one, or a small number of condition(s), but that instead covers those issues that are universally important as a person comes to the end of their life. Such a measure has recently been developed within the capability approach (anonymous3), focusing on the opportunity that a person has to achieve a good death (anonymous1, anonymous6).

The capability approach was developed in the works of Amartya Sen (Sen, 1987, 1993) and Martha Nussbaum (Nussbaum, 2003), and has acquired increasing traction amongst extra-welfarist health economists who want to go beyond a focus just on the health state of the patient (Smith, Lorgelly, Al-Janabi, Venkatapurum, & Coast, 2012). Indeed, there are now a number of index measures of capability developed by health economists, both for generic use (Al-Janabi, Flynn, & Coast, 2012; Grewal et al., 2006) and for use with specific populations, including those suffering chronic pain (Kinghorn, Robinson, & Smith, 2014), receiving public health interventions (Lorgelly, Lorimer, Fenwick, Briggs, & Anand, 2015) and with mental health conditions (Simon et al., 2013). Whilst for some of these measures there has been no attempt to value across different capabilities, the ICECAP suite of measures has used best-worst scaling as a means of attempting to assess the relative value of

the attributes contained in the measures. Such valuation is generally seen as important if the measures are to be used in economic evaluation of interventions.

There are a number of challenging issues in conducting valuations for capability measures, as well as in valuing end of life. In particular, there are important questions about whose values should be used (de Wit, Busschbach, & De Charro, 2000; Dolan, 1999; Ubel, Loewenstein, & Jepson, 2003). In a well-rehearsed argument, extra-welfarist health economists often favour values from the general population, who bear the opportunity cost of decisions about the allocation of health resources (Stamuli, 2011). From a capability perspective too, there may be good reasons for obtaining values from those who are not affected directly by a condition (Coast et al., 2008; Qizilbash, 2006). The values of those in poor health may be subject to adaptation to that condition; with adaptation, values may be higher than they would otherwise be, limiting the scope for change as a result of intervention. This is seen as unfair within the capability approach and, indeed, is a foundation of much of Sen's work (Sen, 1982, 1992) although the reality may be more complex, with some evidence that people focus on the aspects of wellbeing in which they are impaired (Flynn & Huynh, 2015). The use of values from the general population may also align with a Rawlsian 'veil of ignorance' perspective (Rawls, 1971), in that these general population values are generated from those who do not yet know the nature of their own end of life. For those researching in end of life care, however, there is a somewhat different perspective, with an appreciation that end of life is a unique experience that those in the general population may not fully appreciate until they reach this point in their own life (as, of course, everyone does), which may lead to a desire to obtain and use values from those already experiencing the end of life (anonymous1). From this perspective, however, there is also concern that completion of quite hypothetical tasks may be challenging for those at end of life, both physically and emotionally. They may also be costly, requiring sensitive and intensive recruitment, face to face interviews and the provision of health professional support.

A further issue associated with valuing outcomes from health and care, is the question of attribute interaction effects; for example, it may be that having a say in decisions is more important to respondents if they are less able to make adequate preparations for death. Previous attempts to value

capability measures using best-worst scaling have not attempted to estimate interaction effects (Coast et al., 2008; Flynn et al., 2015). Such exploration does, however, require changes to the design of the valuation task beyond the use of a best-worst scaling task alone as this is not designed to estimate interaction effects.

This paper reports research that aims to generate information about values for use in a capability measure of end of life care, the ICECAP Supportive Care Measure (ICECAP-SCM). The measure was generated through qualitative interviews with those at varying stages on the end of life trajectory, including older members of the general population, older people resident in nursing homes and older people accessing palliative care services (anonymous2, anonymous3). Interviews aimed to determine what was important to these people in terms of the care they would receive at end of life, and generally began by asking them about deaths they had experienced in people close to them. This qualitative research, conducted across two rounds of interviews, generated a measure with seven attributes: *choice* (having a say/being able to make decisions about my life and care), *love and affection* (being able to be with people who care about me), freedom from *physical suffering* (experiencing significant physical discomfort), freedom from *emotional suffering* (experiencing emotional suffering), *dignity* (being able to maintain my dignity and self-respect), *support* (being able to have the help and support that I need), and *preparation* (being prepared/having had the opportunity to make the preparations I want to make) (anonymous3). Each attribute has four levels ranging from having none of an attribute, to having it most of the time.

This paper begins by explaining the methods used within the valuation exercise. This includes explanation of the objectives for the valuation task (where a specific aim was to generate a design that would enable direct comparison of valuation responses from those in the general population and those at the end of life), as well as the experimental design and the processes for data collection and analysis. The results are then provided, relating to the generation of two alternative tariffs for use with the measure, one which allows for interactions and the other of which is based only on main effects. The paper ends with a discussion of the methods and results, and explores the policy implications of the research.

ACCEPTED MANUSCRIPT

METHODS

Valuation task

The aim of the valuation task was to obtain as much information as possible from the general population, whilst also nesting within the main valuation task a smaller and less complex task that would provide comparable information if used with individuals at the end of life at a later date. The process by which this nested task was developed is explored in detail elsewhere (anonymous4), but the work was done through extensive piloting to identify "those aspects of the task that would provide comparable data but also be potentially feasible for future use with a vulnerable population group who would find a long and repetitive task too burdensome" (anonymous4).

The valuation task was a stated choice experiment comprising a standard discrete choice experiment with a nested case 2 (profile) best-worst scaling task (Flynn, 2010b; Flynn, Louviere, Peters, & Coast, 2007). Stated choice experiments refer to the quantitative valuation method which collects an individual's stated behaviour (choice) in a hypothetical setting to assess how individuals value goods and services (Louviere, Hensher and Swait, 2000). Stated choice experiments have become an increasingly used quantitative technique in health economics to address a wide range of policy questions including valuing patient preferences and health outcomes, to developing priority setting frameworks and clinical decision making (see de Bekker-Grob et al., 2010 for a review). The survey respondent observes hypothetical, but realistic, scenarios (or alternative possible specifications of a good or service). The individual's decision (which may be a single choice or most and/or least preferred options in each set, ranking or quantities) is elicited across repeated hypothetical scenarios that are systematically varied via an experimental design. The response data elicited is estimated assuming the random utility theory framework.

In the current study, respondents were presented with a set of hypothetical scenarios with each question divided into two phases. First, respondents were shown an initial scenario comprising a profile containing each attribute at a specific level. An example screenshot shows a scenario with

which respondents were presented (Figure 1); here, *Physical suffering* and *Love and affection* are at the lowest level (level 1), *Preparation* is at the second level, *Emotional suffering* is at the third level, and *Support*, *Choice*, and *Dignity* are at the highest level (level 4). For each of these initial scenarios, respondents were asked to choose which aspect (attribute) they considered to be most acceptable (best) and which they considered least acceptable (worst). For the example shown, a respondent might choose the high level of *Dignity* as the most acceptable aspect of the profile and the low level of *Love and affection* as the least acceptable aspect. This comprised the best-worst profile task. After each initial scenario was presented, respondents were then asked to express a preference between the scenario just considered and a 'middling' state (Figure 2) which could only contain attributes at the second level (typically, expressed on the questionnaire as 'some of the time') and the third level (typically, 'a little of the time'), a design previously used to investigate preferences for end of life care among those aged 50 or over (Flynn, Bilger, Malhotra, & Finkelstein, 2016). Respondents were forced to choose between one profile or the other (that is, there was no 'opt out'), on the grounds that it is not feasible to opt out of the end of life entirely. This second element provided the discrete choice data.

In the binary DCE task, individuals make choices among alternatives across attributes or profiles whereas with the best-worst task individuals make within-profile choices (selecting what is best and worst about an option). The appeal of the best-worst task is in its simplicity and potential greater feasibility in eliciting choices from respondents in vulnerable populations (Flynn et al., 2013; Coast et al., 2008), whereas the binary DCE task provides a closer representation of standard stated choice practice and structure. Due to the different contexts in which choice data is elicited, responses elicited may differ and so it is important to assess whether pooling of the data is appropriate.

< Figure 1>

<Figure 2>

ACCEPTED MANUSCRIPT

Experimental design

The ICECAP-SCM contains seven attributes, each with four levels, and so it was clear that respondents could not be expected to consider all 16,384 possible scenarios. Therefore the number was reduced using a Bayesian D-efficient (Bayesian D-error, 0.1742) design using the software Ngene (www.choice-metrics.com). D-efficient designs are constructed by selecting the set of scenarios that minimise the elements of the asymptotic variance-covariance matrix in maximum-likelihood estimation around the set of prior parameters for discrete choice models. By minimizing the elements of the variance–covariance matrix of the model, the statistical efficiency is maximized around the priors (see Rose and Bliemer, 2009 for a review of such designs and Johnson et al, 2013 for constructing designs in health). The final design for this study minimised the variance-covariance of the DCE task using fixed prior estimates obtained from the pilot study (Coast et al., 2016). Due to the small sample size in the pilot study (N=100), the priors were estimated as linear effects (rather than providing estimates for every attribute level) which allowed for more parsimonious estimation of both main and two-way interaction effects (totalling 28 estimated parameters) (see Coast et al., 2016 for details). Given that attribute levels are ordinal in that each of its levels represent a further decrement or impairment in capability, a linear relationship appeared to be a reasonable assumption to use for the initial priors. The priors were assumed to be normally distributed to account for the analyst's uncertainty as to the true population value. Adopting these assumed priors, the final design permitted estimation of each attribute level and linear two-way interaction effects, resulting in the potential estimation of 42 parameters.

Each respondent completed 16 sets of scenarios. The first eight of these were common to all respondents and represented a core set from which analysis of heterogeneity could be conducted; the second eight were from one of five blocks within the design. Blocks were generated in a way that minimised the correlation structure between the blocking variable and the design attributes. This has the effect of minimising the potential confound between the individual and the block that they are assigned to. Respondents were randomly allocated at the point of entering the survey to receive one of these five blocks. The design was optimised for the DCE task which allowed for interactions. The

element of this task which could potentially be used to provide comparable information from those at the end of life at a later date was the best-worst scaling element of the first eight (core) scenarios.

Overall survey design and conduct

The valuation exercise was accompanied by a number of general socio-demographic questions and a general wellbeing measure (ICECAP-A) (Al-Janabi et al., 2012) as well as questions specifically concerning the experience of the individual in relation to end of life care, including questions about personal life-limiting illness, caring roles, bereavement and medical or nursing experience.

The research was reviewed and approved by the University of Birmingham's Science, Technology, Engineering and Mathematics Ethical Review Committee [ERN_11-1296].

Piloting

The survey was piloted through the same online panel as was used in the main study. Three online pilots were conducted between September 2012 and March 2013 with a total of 406 respondents. These pilots informed a number of design issues, including the experimental design (based on priors obtained from the pilot study), and some wording issues (anonymous4).

Sampling

The sample for the survey was the UK adult general population. The survey was conducted through online panel provider Pureprofile, which at the time had a UK panel of over 600,000 respondents. The sampling was targeted to be representative of the UK general population on the key socio-demographic characteristics of age and gender.

Values for previous ICECAP measures have been based on best-worst scaling alone and have used face-to-face interviews. To determine the appropriate sample size for a valuation exercise, prior

knowledge about the preference structure (utility function) and coefficients are used as priors in designing efficient experimental designs. While the pilot studies provided some indication of the general direction and average coefficients from a conditional logit model to use as priors, prior knowledge of heterogeneity in the preferences (e.g. preference segments), scale or decision strategies and processes (e.g. compensatory versus non-compensatory) would also be needed to inform the design and sample size. For this study, prior knowledge of the systematic variability was not known and it was particularly desirable to be able to account fully for any potential preference heterogeneity and/or other sources of heterogeneity, given the difficult subject matter and the possibility that diverse groups may have very different values. Such differences in value may relate in part, but not exclusively, to individuals' experiences of end of life, both personally and among those they are close to. Given the ability to recruit larger sample sizes through an online panel, a large sample size of 6000 individuals was proposed.

Once the 6000 sample size had been reached, sampling continued until all block sizes were balanced (i.e. equal in number).

Analysis

The broad analytic strategy was to begin with analysis of the best-worst data, followed by analysis of the DCE data. Consideration was then given to the acceptability of pooling across these two data sets, to enable analysis on the combined data set.

Analysis of best-worst scaling data

The analysis of best-worst scaling is rooted in random utility theory (McFadden, 1974; Thurstone, 1927); the estimated values are a function of choice frequencies and the choice of an attribute at a particular level indicates how much the individual values each level of each attribute over the other (Louviere, Flynn, & Marley, 2015). The appropriate econometric model is a limited dependent variable model – generally, as a first step, a conditional logit (also known as a multinomial logit, or

MNL) model in which choice options are simply attribute levels rather than complete states. The "worst" choice data are appended to "best" choice data with all independent variables taking a sign change (reflecting that "worst" from options with utilities of, say, 3,5,8,9 is observationally equivalent to "best" from options with utilities -3,-5,-8,-9).

A sequential psychological model was assumed for the analysis of best-worst choices, leading to a partial ranking of the attribute levels chosen (Flynn, Louviere, Peters, & Coast, 2010) and operationalised with the conditional logit model. A latent class cluster analysis was run using the best minus worst scores to provide an initial exploration of the preference heterogeneity in the data (Flynn, 2010a; Flynn et al., 2015; Marley & Louviere, 2005).

A series of scale-adjusted conditional logit latent class analyses were conducted using the combined best-worst data to separately identify preference and scale heterogeneity independently of covariates (Flynn et al., 2010). This was conducted to enable the identification of clusters representing "types" of respondent who differed in their relative preferences for the seven attributes of ICECAP-SCM. Latent Gold Choice (with syntax module) software was used. The behavioural model assumed was a logit model, and the preference distribution a discrete finite mixture of logit models assumed to comprise latent classes of respondents with the same preference part-worth utilities and/or scale. Similar assumptions were made in the analysis of ICECAP-A data (Flynn et al., 2015). Both the Bayesian Information Criterion (BIC) and the stability of solutions (tested through using different starting seeds) were used to guide model selection and choose the optimal model. Models were then run with covariates (sociodemographic variables) to allow for covariate specific heterogeneity, that may manifest either through identifying class membership or directly via the utility function. Univariable analyses were conducted to test for possible predictors and then a multivariable analysis was performed (Patel, Peters, Murphy, & Team, 2005; Peters, Somerset, Campbell, & Sharp, 2003). Possible predictors included general demographic variables such as age, gender and employment status, as well as a range of variables capturing experiences that are expected to potentially influence preferences for end of life care (such as bereavement, caring for someone with a long-term illness and diagnosis of a long-term illness).

Pooling of best-worst and DCE data and combined analysis

Two exploratory steps were undertaken before any formal testing or pooling of data. Initially, multinomial logit models were estimated from both the best-worst and DCE data. For both, the model assumed was a main effects MNL model, enabling comparability between these two datasets to be assessed. Coefficients from the two models were then plotted to determine whether the preferences expressed through the two methods were similar (proportional) (Hensher, Louviere, & Swait, 1999; Swait & Bernardino, 2000). This provides an initial indication as to whether pooling of data is likely to be possible, that is, that respondents are using similar cognitive processes when completing each task, albeit with a change in the decision context. Secondly, a further set of scale-adjusted latent class analyses were run on the DCE data, with covariates to identify class membership. Methods used were consistent with those reported for the best-worst scaling data. The scale-adjusted latent class analyses for both the best-worst data and the DCE data were then considered in terms of whether they produced similar results for the numbers of classes, and the composition of these classes.

Following this stage, formal testing for the pooling of the data was conducted. This followed the Swait & Louviere (1993) approach to pooling different sources of data. A Chow test for pooling data (Swait & Louviere, 1993) was used. The test statistic compares the sum of the log likelihood from the MNL models for each choice context to a log likelihood function from a heteroscedastic conditional (multinomial) logit model that adjusts for scale differences across the data sources (Swait, 2006). The hypothesis was homogeneity in preferences across the best-worst and the DCE tasks, whilst allowing for variance scale differences across these two datasets. Passing this test at the 5% significance level indicated that it would be appropriate to combine the datasets as long as the different variance scales were accounted for.

The combined data were analysed with a heteroscedastic conditional (multinomial) logit model adjusting for continuous scale run by code in GAUSS version 14. Datasets were combined by

stacking the best-worst data on top of the DCE data, such that main effects were combined into the same variables, whereas interaction effects were DCE specific. We also allowed for indicators to account for scale differences between best data, worst data and DCE choice data; these indicators were effects coded. Attributes were coded as continuous linear decrements from the top level of that attribute (that is, the highest attribute level was coded as zero, and each level below that was coded as being -1 and -2 and -3). The decision to run the model in this way related to the development of the design following analysis of the pilot data. This coding also made it easier to run a model with both main effects and interaction effects, by reducing the number of interactions from 84 to a manageable number of 21. The final indirect utility function to be estimated for decision maker *n* is:

$$U_n = exp(\mu_{BWdata} + \mu_{Best})\,[Const + \Sigma_i\, ASC_i + \Sigma_i \beta_i X_{in} + \Sigma_i \Sigma_{j \neq i} \gamma_{ij} X_{in} X_{jn} + \Sigma_i \Sigma_l \delta_{ln} Z_{ln} ASC_i + \Sigma_i \Sigma_l \delta_{ln} Z_{ln} X_i]$$

EQUATION (1)

Where *Const* is the constant for the middling alternative in the DCE task, ASC_i are the alternative specific constants for attribute *i*, which represents on average the attribute impact upon utility. The variable X_i represents the linear decrements for attribute *i* of k=1,…,7 attributes, and X_i X_j are the interaction terms between decrement values for attributes *i* and *j*. Covariate *l* for individual *n* is denoted as Z_{ln}, where its interactions with the alternative specific constants and with attribute *i* is $Z_{ln}ASC_i$ and $Z_{ln}X_i$, respectively. The parameters β_i, γ_{ij} and δ_{ln} are to be estimated. The scale factor is captured in the exponentiated expression in the equation, in which μ_{BWdata} relates the scale parameter for best-worst data relative to the DCE data is and μ_{Best} the scale parameter for the best data relative to the worst data.

Guided by the previous analyses of the best-worst data and DCE data, the team selected covariates that would be most important to include in the analysis (given that, with the large amount of data it was not feasible to run the model with all covariates). Each of these was tested independently and then introduced to the final model.

Development of tariffs

As with the previous ICECAP measures, final tariffs were developed from the initial best-worst analysis. However, here a second set of tariffs was also developed from the pooled data. Interpretation of these two tariffs is discussed at the end of the paper. Following Flynn, Huynh et al. (2015), the heterogeneity-adjusted population level tariffs were calculated as follows. First, a linear transformation was applied to the utilities to generate tariffs that were anchored at both 'full capability' and 'no capability' where these states are represented by 4444444 (highest capability state on each attribute) given a value of one and 1111111 (lowest capability state on each attribute) was given a value of zero, respectively. This is achieved by, first, subtracting one seventh of the value of state 1111111 from all attributes and then dividing by the index value for 4444444. This rescaling ensures that the 'no capability' state sums to zero, that 'full capability' state sum to one while still maintaining relative differences between attributes and attribute levels. (See Coast et al., 2008; Flynn et al., 2015).

Second, the average population tariff was calculated, by taking the weighted mean of the tariffs across classes. The class membership posteriors were used as weights and account for preference class membership and 'net out' differences in scale. Second, to ensure representativeness, population weights were applied to match UK 2011 Census population statistics (England, Wales, Scotland and Northern Ireland). Similar to Flynn and Huynh (2015), the posterior class membership predictions for each individual in the sample is used to identify the sample characteristics for each class and the tariffs from each class is then re-weighted by applying population weights matching the population statistics. Demographic variables selected to apply population weights are informed by large differences detected from initial comparison of descriptive statistics against the population: age, gender and employment.

ACCEPTED MANUSCRIPT

RESULTS

In total, 40,831 individuals were approached to take part from the Pureprofile UK online panel during June 2013. From these there was an initial response from 8,336 individuals of whom 382 (4.6%) were screened out because they were too young or did not consent and 1,796 (21.5%) did not fully complete the survey. A total of 6,158 (73.9%) completed, but the final 138 responses were dropped to ensure balance across blocks, giving a final sample size of 6,020 (14.7% of those approached) of UK adults aged 18 and over. Analyses are reported for these 6,020 responses; it seemed appropriate to remove the final responses that exceeded the balance quota for each block, but it should be noted that the main findings appear robust to removing different respondents in the sample to maintain block balance.

The median time spent on completion of the survey was 17 minutes. Table 1 shows the characteristics of respondents and Table 2 the responses to the ICECAP-A capability wellbeing questionnaire. The sample of respondents is roughly representative of the UK population in terms of gender. While age in the sample is similarly increasing across age bands compared to the population, there is a larger proportion (19.6%) of respondents in the sample aged 55-64 years compared to the UK population (7.7%). Of particular interest is that around 10% of this general population sample were suffering from a life-limiting illness and around 40% had experienced a close bereavement in the previous two years. As with an earlier face-to-face general population survey using ICECAP-A (Al-Janabi et al., 2013), the sample here had modal responses to the ICECAP-A that were the highest or second highest level of capability for each of the five domains. Many individuals did however have lower levels of capability (little or none) on each of the five domains. This ranged from 8% of individuals for autonomy to 21% of individuals for achievement.

<Table 1>

<Table 2>

Results from best-worst scaling data and derivation of best-worst tariff

The intial MNL model of the best-worst scaling data suggested that *Dignity* and *Support* were highly valued (covered a relatively large proportion of the utility space), and that freedom from *Emotional suffering* and *Preparation* were the least valued of the attributes. These results, however, are crude in that they do not capture any potential preference heterogeneity within the data. Detailed reporting of the data from the initial MNL model is provided later in the paper in relation to pooling the best-worst and DCE data.

The cluster analysis identified four clusters, the smallest of which (n=900) comprised individuals with no strong preferences. Two clusters were both interested in *Dignity* and minimising *Physical suffering*, but the largest (n=2370) additionally had strong concerns for *Support*, and *Love and affection*, whilst the smaller (n=1380) did not have these additional strong concerns. The remaining cluster (n=1370) whilst also valuing *Dignity* and *Love and affection*, had a strong concern for *Choice*.

In the latent class conditional logit analysis, four preference classes and two scale classes were identified. The final model was also adjusted for covariates and these estimates are shown in Table 3. It is also helpful to visually represent the findings to aid interpretation and this is done in Figures 3(a) and 3(b). The two scale classes suggested that one class was approximately twice as consistent as the other.

<Table 3>

The four preference classes were similar to those of the cluster analysis. Class 1 (n=1419, 23.6% unweighted; 27.3% weighted) comprises individuals who had weak preferences that, if anything, focused largely on *Support* and *Emotional suffering* (see Figure 3(a)), and where there was little distinction between levels of attributes (as shown in Figure 3(b)). For this class, the levels for Physical suffering and Emotional suffering were restricted to zero in response to the statistically insignificant results from a joint F-test across the levels for these attributes. Those in this class were significantly more likely to be younger, male, employed full-time, religious and with a life-limiting illness. Class 2 (n=1639, 27.2% unweighted; 29.3% weighted), which was characterised by

individuals who were older, female, not employed full-time and had been recently bereaved, had its strongest concern for *Dignity*, but also gave high value to *Choice* and *Support* and had some overall concern with both types of suffering. Class 3 (n=1685, 28% unweighted; 25.2% weighted) had strongest values for *Love and affection*, but also gave high values to *Support* and *Dignity*, and was characterised by being younger, being female, looking after the home full-time but not acting as an informal carer. Finally, Class 4 (n=1277, 21.2% unweighted; 18.2% weighted) particularly strongly valued minimising *Physical suffering*, with high value also given to *Support*. This final group was characterised by those who were older, male, not religious and on long term sick-leave, but without experience of death and dying (no life-limiting illness and not recently bereaved). Across all classes, *Emotional suffering* was not highly valued, nor were individuals particularly sensitive to differences in the levels (Figures 3(a) and 3(b)). This contrasted with *Preparation* which was more highly valued across classes two to four (Figure 3(a)), but where, again respondents were not particularly sensitive to differences in levels (Figure 3(b)).

<Figure 3(a)>

<Figure 3(b)>

From these data, the best-worst tariff for ICECAP-SCM was also derived and is shown in Figure 4. Population weights were applied to the tariffs presented based on employment, gender and age from UK Census 2011, with the ratio of population to sample proportions ranging from 0 (Male, 65 years and over, unemployed) to 7.5 (Male, 65 years and over with 'other' employment). The final tariffs for each of the four capability levels for that attribute (4 indicating full capability) are presented in the table below Figure 4each attribute in Figure 4. For example, with this tariff, the value for state 2222222 is the sum of tariffs at level 2 across all attributes, equalling to 0.373. The tariff support is around 24% of the capability space, love and affection 18%, Dignity 17%, prepared 14%, choice 13%, physical suffering 10% and emotional suffering 4%suggests that all of the seven attributes of the ICECAP-SCM make significant contributions to an individual's capability at the end of life, with *Support* accounting for 24% of the capability space, *Love and affection* and *Dignity* each accounting for around 17-18%, *Choice* and *Physical suffering*, each accounting for around 13-14%, *Preparation*

accounting for 10% and *Emotional suffering* having a much smaller contribution of only around 4% of the capability space. Across all the attributes, the differences between shifts at the lower and higher levels were statistically similar (that is, the differences in value between having no capability and some capability is similar to that between being able to have some capability and a lot of capability). As with the other ICECAP tariffs, which are also based on best-worst scaling data, the tariff is additive.

<Figure 4>

Assessment of whether best-worst and DCE data could be appropriately pooled

Exploratory analyses were conducted to assess whether the best-worst and DCE data could be appropriately pooled before a formal pooling test was conducted. A plot of the coefficients obtained from multinomial logit models obtained from both the best-worst and binary choice DCE data indicated responses to the best-worst task were around 3.5 (=1/0.296) times more consistent than those from the DCE task. These results were largely consistent across each individual attribute (results available upon request).

To explore whether similar patterns of preference heterogeneity arise from the best-worst data and binary DCE data, scale-adjusted latent class analysis produced four preference classes for both the best-worst data and for the DCE data. As these analyses were conducted primarily with the aim of informing the data pooling, they are described only briefly here. Two of the classes from the DCE data seemed to equate particularly well to the best-worst scaling classes, with one class very valuing *Love and affection* and *Support*, and a second class valuing the minimisation of *Physical suffering*. The other two classes were more loosely related. It should be noted that a perfect correspondence across the two data sources would not be expected, partly because of the different contexts for the two tasks and partly because the interactions are not interpreted for the DCE data.

Given the apparent similarities in the data sources from these two exploratory analyses, the formal test for pooling data sources was conducted. The test statistic compares the sum of the log likelihood from

the DCE analysis and best-worst analysis, to a log likelihood function from a heteroscedastic conditional logit model that adjusts for scale differences across the data sources. The log likelihood from the analysis of the best-worst task (heteroscedastic conditional logit model that included an indicator for best -worst data in the scale function) was -290,989.16 and for the MNL analysis on DCE data was -57,031.24. The pooled model was a heteroscedastic conditional logit model that included effects coded indicators to account for scale differences between best and worst data, and DCE choice data, which produced a log likelihood of -348,034.41. This resulted in a test statistic of -2(-290,989.16 - 57,031.24 + 348,034.41), equals 28.03, compared to the critical value at the 5% level of 32.67 (21 df). As the test statistic is below this critical level, the data can be appropriately pooled in that the values are the same and the relative scale factor measures the heterogeneity of the error variances across all the data sources.

ACCEPTED MANUSCRIPT

Results from pooled best-worst and DCE data and derivation of combined tariff

The latent class analysis across the pooled data sources resulted in four preference classes reported in Table 4. The latent classes broadly correspond to the classes identified in the best-worst analysis: Pooled Class 1 (29% weighted), was characterised by people who had weaker preferences in general, but focused on *Supportt*, closely corresponds to Class 1 *"Weak preference: Support"* of the best-worst analysis. Closely corresponding to Class 3 "*Love, affection*, *Support* and *Dignity*" of the best-worst analysis, the pooled Class 2 (32% weighted) comprised those who particularly valued *Love and affection* and *Support* but *Dignity* is not as strongly preferred. Pooled Class 3 (19% weighted) comprised those who had a particular preference for minimising *Physical suffering*, whilst Pooled Class 4 (20% weighted) was made up of those individuals who valued *Dignity* and *Choice*. Both classes correspond to Classes 2 and 4 in the best-worst analysis, respectively.

Sixteen out of 21 two-way interactions were statistically significant within the pooled data. The interaction estimates presented in Table 4 are the effects of interactions between attribute decrements (taking on non-positive values 0, -1, -2, -3, representing impairments from full capability) on the values. A negative interaction effect indicates a further decrease in value, while a positive interaction parameter captures additional improvements in value when both attributes are at less than full capability. The results in Table 4 show that simultaneously having some level of physical suffering together with less than full dignity is considered worse for Classes 1, 2 and 3. Interaction effects are better understood when interpreted together with the main effects: For example, while both *Support* and minimising *Physical suffering* are important for Class 3, and any decrements from full capability for each of these attributes have a strong negative effect (positive parameter multiplied by non-positive decrement), this effect is reduced via the interaction term.

Guided by the previous analyses of the best-worst data and DCE data, the covariates age, gender, and being diagnosed with life-limiting illness, bereavement, employment and religiosity were considered in the analysis. From the previous analyses, older people are expected to be strongly associated with valuing *Dignity*, *Choice* and *Support*, and similarly for those that have experienced bereavement. Being diagnosed with life limiting illness or being on long-term sick leave or disability (as captured

by the employment variable) is related to minimising *Physical suffering*. Females were expected to value love and affection more so than men. Associations for religiosity were less clear. The covariates that were significant in the pooled model were age, gender and being diagnosed with life-limiting illness. The covariates were not found to be significant in explaining class membership, but were significant as interactions included directly in the utility function. It is important to re-iterate here that the results from the best-worst analysis (Table 3) and the pooled analysis (Table 4) are not expected to correspond perfectly due partly due to differences in context of the choice task and data structure and the functional form assumed. While it is reassuring to observe some similarities in preference structure across the two analyses, it is not surprising to find differences in the effect of sociodemographics across analyses. In line with expectations, respondents aged over 65 valued *Dignity* and *Support* more highly compared to people under the age of 65, and females valued *Love and affection* more highly than males. Interactions of the main effects with a life limiting illness diagnosis were significant but the effect varied by class and did not match our initial findings.

The best data had a larger scale than the worst data by 4.8, suggesting that respondents were more consistent in identifying what they most valued than what they least valued. In terms of the difference between best-worst data and DCE data, the scale was roughly one between these data sources indicating that respondents did not find the DCE task any harder to do than the best-worst task.

The pooled tariff contains 16,384 states, each calculated using the output from the analysis in combination with equation (1). The value of the state is first calculated based on the raw utility estimates by class and the average population value is calculated by taking the (population) weighted mean of the total tariffs across classes. The tariff is anchored at 4444444 equal to one (full capability at end of life) and 1111111 equal to zero (no capability at the end of life). This is achieved by subtracting from the value of the 'no capability' divided by the difference in value between 'full capability' and 'no capability'. With this tariff the state 2222222 has a value of 0.175 and the state 3333333 a value of 0.508. Interestingly, and different from the other ICECAP measures, with this tariff 15,606 (95.2%) of possible states within the end of life measure are valued at below 0.6. It should be noted that there are, however, some issues with the tariff. Because of interactions, zero is

not the lowest score; there are 78 states that, from this model, produce an unintuitive tariff finding, in that an improvement in one or more attributes from the 'no capability' state results in a lower score than zero. For example, state 1224111 has a value of -0.012, and state 1112111 has a value of -0.027. Because it is not feasible to have negative capability, and it is a tiny proportion of states (0.0048) that are affected, the decision was made to truncate all negative values at zero. Tariff scores for all 16,384 states from this model are available from a look-up table (Online Appendix 1).

<Online Appendix 1, as link to the ICECAP website>

ACCEPTED MANUSCRIPT

DISCUSSION

This research has produced the first valuations for the ICECAP-SCM, and indeed the first valuation set for use with a generic measure of end of life care that has appropriate attributes for use in this particular setting. This is important because it will enable the conduct of economic evaluation utilising measures that capture the aspects of end of life care that are important to people at this time, and give appropriate weight to these different aspects of end of life care. These tariffs can be applied to shifts between states as the result of interventions/services, to provide a quantified estimate of the gain (or sometimes loss) in benefit from those interventions/services; this information can then be combined with cost data to assess efficiency and inform decision making.

The research is also the first to use pooled DCE & best-worst data to generate tariffs for use in health economics, and one of the first exercises using both data sources to also estimate interactions. The research has resulted in two tariffs. The first is based on the best-worst data alone; the second on the pooled data. There are advantages and disadvantages of using each in a policy context, which are discussed below. Additionally, the research used a novel design that means that it will in time be possible to compare directly, results from this exercise with the general population, with results from a population of individuals at the end of life.

The tariffs generated from the pooled DCE and best-worst data contain some states with unintuitive values. Although it is not possible to provide definitive explanations for these oddities, there are a number of plausible explanations. The first is that some of the states that respondents were asked to value may have contained some aspects that they did not find plausible, leading to difficulties for the respondents in providing answers that truly reflected their values. The second is that, given the small number of states that were valued relative to the total number of interactions, the need to extrapolate beyond the data collected may have led to these unintuitive values. The third is that there may be three way interactions that it was not possible to capture within the design and analysis. If there are three way interactions, then these will be confounded with the main effects and the two way interactions that are observed. The differences in scale factor between the two exercises can be

interpreted in a number of ways. Having a larger scale suggests that, for the same amount of effort, it is possible to achieve smaller standard errors, that is, more precise estimates. This could be interpreted to mean that there are stronger preferences or that the task is less cognitively burdensome.

In this exercise, greater linearity was found in the best-worst tariffs than in those for similar valuations for either ICECAP-A (Flynn et al., 2015) or ICECAP-O (Coast et al., 2008). There is no obvious reason for this: it may be related to the choice of levels for this measure, which differ slightly from those of the other two ICECAP measures in that the top level relates to 'most' rather than 'all', and that the levels are expressed in term of time rather than severity; or it may be related to the particular attributes included in the ICECAP-SCM. This, as well as the generation of values for the interactions in the pooled tariff, results in higher numbers of states with low values than in the ICECAP-A and ICECAP-O tariffs. As yet, there are no comparable end of life measures that have attempted to generate values for use of the measures in economic evaluation, although there is work ongoing to generate economic values for the Palliative care Outcomes Scale (POS) in the form of the POS-QOL (http://pos-pal.org/maix/background.php). In time, it will be important to compare the findings here with this work.

This work has both strengths and limitations. The work was carefully designed to simultaneously achieve a number of aims (anonymous4). These included the ability to look at interactions (requiring a relatively complex task) whilst also being able to compare general population values with those at end of life (requiring a relatively simple task). The research was also conducted with a large sample size, enabling robust results to be generated in terms of the analysis of heterogeneity. While the research controls for a number of demographic and socio-economic variables, the final set of covariates in the study did not include direct measures for socio-economic status such as education. It is possible that an individuals' preferences for their end of life care could be shaped by their positioning on the socioeconomic ladder. Socio-economic status was only captured to some extent through occupation, employment and variables relating medical training experiences. Another limitation of the work is that it was conducted using an online panel, and there may be some differences between those members of the public who join such online panels and those who do not.

As discussed by Boyle, Morrison et al (2015), online panel companies such as Pureprofile (used in this study) have developed panels for administration of internet surveys where characteristics are matched to population characteristics. While respondents are randomly selected, survey non-response can introduce representation error. Boyle, Morrison et al (2015) show that most differences were identified to be due to the survey mode and not the survey framing, while other studies suggests that use of online panels has similar validity to face-to-face interview (Mulhern et al., 2013). Nevertheless, we have ameliorated this issue as far as possible by adjusting for national representativeness by employing population weights in the final tariffs.

Clearly this work, which has generated two alternative tariffs, could result in a number of different recommendations for public policy. In general, when generating tariffs, analysts have done considerable work, but presented one final model as the model for use in policy decisions. Here, we present two alternative tariffs based on different data and methods. We present both in the interests of being transparent about the different options, and suggest that, to some extent, choice of the tariff depends on the objectives of policy makers. The tariff based on the best-worst data alone provides a simple off-the-shelf scoring that can be easily applied to the tick box answers that people give in their scoring. The tariff can be easily understood given its simple additive nature; it is easy to interpret and is well-behaved. It also retains consistency with the nature of the tariffs for the two previous ICECAP exercises and the context in which the task is completed is more closely aligned to Sen's notion of values, where individuals are not asked to trade one attribute for another, but merely to express values (Coast et al., 2008). Negatives for this tariff are that it is based only on the main effects, assuming that effects are independent and additive, and that it ignores the interactions found in the more complex analysis. Additionally, the statistical design for the experiment was powered for the DCE allowing for the estimation of two-way interactions, rather than for the best worst scaling task.

The tariff arising from the pooled data, on the other hand, gives this information about two-way interactions and utilises the design fully; it also draws on the sort of context of trading between one option and another that is generally preferred by economists in valuation studies (Parkin & Devlin, 2006). The tariff, however, gives unintuitive results for some states and indeed, in a few situations

has unpalatable implications. Even truncating all negative values at zero (on the grounds that negative capability is not feasible) leaves some oddities in place. For example, use of the pooled tariff suggests that if there is a shift from no capability on any attribute to one with no capability on any attribute apart from support, even where support is at the highest level, there will not be any resulting improvement in value. The unpalatable implication is that in a world focusing on cost-effectiveness and because it would not improve benefit to the patient, no support need be offered to patients in this group with severe limitations on capability. One important issue in choosing between the two tariffs may be ascertaining whether the problematic states in the pooled tariff arise much, or indeed at all, in practice and thus the extent to which this is a real problem or a theoretical one.

This work suggests a number of avenues for further research. One is to ascertain the extent to which the different tariffs (value sets) generate different decisions in empirical research. Experience in using the tariffs may also generate further issues of comparability between them. A second clear area of further research is to carry out the planned exercise with individuals at the end of life, to determine the extent to which values of those at the end of life coincide with those in the general population – or subgroups of the general population - and thus to further inform the choice of tariff. This is important given the differing perspectives around the issue of whose choices should be used. If values from the two perspectives are aligned then choice of tariff is less problematic; if they are not, then decision makers choice about which tariff to use are likely to be affected by their particular perspective, with decision makers in organisations such as the UK's National Institute of Health and Clinical Excellence likely to prefer a general population set of values in alignment with their more general approach, and charitable decision makers such as hospices possibly preferring to prioritise the values of those at end of life.Indeed, one interesting finding from the best-worst scaling results was that there seemed to be greater value attached to freedom from physical suffering among those without such experience. Further exploration of this issue might be appropriate, as it does suggest some shift in values amongst those with experience of end of life. Given that patients may be very unwell and may find the sorts of questions asked in these valuation exercises very tiring, a simplified but comparable, best -worst scaling task is proposed. Feasibility of conducting the simpler exercise with those at end

of life is currently being tested, but previous successful discrete choice experiments amongst end of life patients (Douglas, Normand, Higginson, & Goodwin, 2005; Finkelstein, Bilger, Flynn, & Malhotra, 2015; Malhotra, Farooqui, Kanesvaran, Bilger, & Finkelstein, 2015; Morton et al., 2012) give grounds for optimism that this research will be possible to conduct.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

CONCLUSION

This work has conducted an extensive valuation exercise for the ICECAP Supportive Care Measure with a large number of members of the general public. The experiment was carefully designed to facilitate future comparison with studies among those experiencing the end of life. The findings of the work with the general public have produced two alternative tariffs, one of which is unusual amongst economic valuation studies in that it fully accounts for two way interactions; the other is simpler and has more intuitive implications. Choice of tariff for empirical work should depend in part on the views of decision maker, taking account of the advantages and disadvantages of each tariff, and in part on future empirical exploration of the implications of using the tariffs in practice.

ACCEPTED MANUSCRIPT

REFERENCES

Anonymous 1-6

Al-Janabi, H., Flynn, T. N., & Coast, J. (2012). Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research, 21*, 167-176.

Al-Janabi, H., Peters, T. J., Brazier, J., Bryan, S., Flynn, T. N., Clemens, S., Moody, A., & Coast, J. (2013). An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research, 22*, 1831-1840.

Boyle, K. J., Morrison, M., MacDonald, D. H., Duncan, R., & Rose, J. (2016). Investigating internet and mail implementation of stated-preference surveys while controlling for differences in sample frames. *Environmental and Resource Economics*, 64(3), 401-419.

Canada, H. (2007). *Canadian strategy on palliative and end-of-life care. Final report of the coordinating committee. December 2002 to March 2007*. Ottawa: Health Canada.

Coast, J., Flynn, T. N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J. J., & Peters, T. J. (2008). Valuing the ICECAP capability index for older people. *Social Science and Medicine, 67*, 874-882.

de Wit, G. A., Busschbach, J. J. V., & De Charro, F. T. (2000). Sensitivity and perspective in the valuation of health status: whose values count? *Health Economics, 9*, 109-126.

Department of Health. (2008). *End of life care strategy. Promoting high quality care for all adults at the end of life.* London: Department of Health.

Department of Health and Ageing. (2010). *Supporting Australians to live well at the end of life. National palliative care strategy 2010*. Canberra: Department of Health and Ageing.

Dolan, P. (1999). Whose preferences count? *Medical Decision Making, 19*, 482-486.

Douglas, H., Normand, C. E., Higginson, I. J., & Goodwin, D. E. (2005). A new approach to eliciting patients' preferences for palliative day care: the choice experiment method. *Journal of Pain and Symptom Management, 29*, 435-445.

Finkelstein, E. A., Bilger, M., Flynn, T. N., & Malhotra, C. (2015). Preferences for end-of-life care among community-dwelling older adults and patients with advanced cancer: a discrete choice experiment. *Health Policy, 119*, 1482-1489.

Flynn, T. N. (2010a). Using conjoint analysis and choice experiments to estimate quality adjusted life year values: issues to consider. *Pharmacoeconomics, 28*, 711-722.

Flynn, T. N. (2010b). Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Review of Pharmacoeconomics & Outcomes Research, 10*(3), 259-267.

Flynn, T. N., Bilger, M., Malhotra, C., & Finkelstein, E. A. (2016). Are efficient designs used in discrete choice experiments too difficult for some respondents? A case study eliciting preferences for end-of-life care. *Pharmacoeconomics, 34*(3), 273-284.

Flynn, T. N., & Huynh, E. (2015). BWS profile case application: preferences for quality of life in Australia. In J. J. Louviere, T. N. Flynn & A. A. J. Marley (Eds.), *Best-Worst Scaling. Theory, methods and applications* (pp. 240-262). Cambridge: Cambridge University Press.

Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Economics, 24*(3), 258-269.

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best-worst scaling. What it can do for health care research and how to do it. *Journal of Health Economics, 26*, 171-189.

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2010). Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. *Social Science and Medicine, 70*, 1957-1965.

Gawande, A. (2014). *Being mortal. Illness, medicine and what matters in the end.* London: Profile Books Ltd; Wellcome Collection.

Gawande, A. (2016). Quantity and quality of life. Duties of care in life-limiting illness. *JAMA, 315*(3), 267-269.

Grewal, I., Lewis, J., Flynn, T. N., Brown, J., Bond, J., & Coast, J. (2006). Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science and Medicine, 62*, 1891-1901.

Hensher, D., Louviere, J., & Swait, J. (1999). Combining sources of preference data. *Journal of Econometrics, 89*, 197-221.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B., Kanninen,B., & Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health*, 16(1), 3-13.

Kinghorn, P., Robinson, A., & Smith, R. D. (2014). Developing a capability-based questionnaire for assessing well-being in patients with chronic pain. . *Social Indicators Research, 120*, 897-916.

Lorgelly, P. K., Lorimer, K., Fenwick, E. A. L., Briggs, A. H., & Anand, P. (2015). Operationalising the capability approach as an outcome measure in public health: the development of the OCAP-18. *Social Science and Medicine, 142*, 68-81.

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling. Theory, methods and applications.* Cambridge: Cambridge University Press.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated Choice Methods: Analysis and Application* (1st ed.). Cambridge; United Kingdom: Cambridge University Press.

Malhotra, C., Farooqui, M. A., Kanesvaran, R., Bilger, M., & Finkelstein, E. A. (2015). Comparison of preferences for end-of-life care among patients with advanced cancer and their caregivers: a discrete choice experiment. *Palliative Medicine, Online first*. doi: 10.1177/0269216315578803

Marley, A. A. J., & Louviere, J. J. (2005). Some probabilistic models of best, worst and best-worst choices. *Journal of Mathematical Psychology, 49*, 464-480.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105-142). New York: Academic Press. (Reprinted from: Not in File).

Morton, R. L., Snelling, P., Webster, A. C., Rose, J., Masterson, A., Johnson, D. W., & Howard, K. (2012). Factors influencing patient choice of dialysis versus conservative care to treat end-stage kidney disease. *Canadian Medical Association Journal, 184*(5), E277-E283.

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., & Tsuchiya, A. (2013). Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value in Health, 16*, 104-113.

Normand, C. (2009). Measuring outcomes in palliative care: Limitations of QALYs and the Road to PalYs. *Journal of Pain and Symptom Management, 38*, 27-31.

Normand, C. (2012). Setting priorities in and for end-of-life care: challenges to the application of economic evaluation. *Health Economics, Policy and Law, 7*(4), 431-439.

Nussbaum, M. C. (2003). Capabilities as fundamental entitlements: Sen and social justice. *Feminist Economics, 9*(2-3), 33-59.

Parkin, D., & Devlin, N. (2006). Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics, 15*, 653-664.

Parliamentary Assembly. (2009). *Resolution 1649 (2009). Palliative care: a model for innovative health and social policies*: Council of Europe.

Patel, R. R., Peters, T. J., Murphy, D. J., & Team, A. S. (2005). Prenatal risk factors for Caesarian section. Analysis of the ALSPAC cohort of 12944 women in England. *International Journal of Epidemiology, 34*, 353-367.

Peters, T. J., Somerset, M., Campbell, R., & Sharp, D. J. (2003). Variables associated with attendance at, and the perceived helpfulness of, meetings for people with multiple sclerosis. *Health and Social Care in the Community, 11*, 19-26.

Qizilbash, M. (2006). Well-being, adaptation and human limitations. *Royal Institute of Philosophy Supplement, 59*, 83-110.

Rawls, J. (1999). *A theory of justice. Revised edition.* Oxford. Oxford University Press.

Rose, J. M., & Bliemer, M. C. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews, 29*(5), 587-617.

Sen, A. (1982). *Choice, welfare and measurement*. Cambridge, MA: Harvard University Press.

Sen, A. (1987). The standard of living: lecture II, lives and capabilities. In A. Sen (Ed.), *The standard of living* (pp. 20-36). Cambridge: Cambridge University Press. (Reprinted from: Not in File).

Sen, A. (1992). *Inequality reexamined*. New York: Russell Sage Foundation.

Sen, A. (1993). Capability and well-being. In M. C. Nussbaum (Ed.), *The quality of life*. Oxford: Clarendon Press. (Reprinted from: Not in File).

Simon, J., Anand, P., Gray, A., Rugkasa, J., Yeeles, K., & Burns, T. (2013). Operationalising the capability approach for outcome measurement in mental health research. *Social Science and Medicine, 98*, 187-196.

Smith, R. D., Lorgelly, P., Al-Janabi, H., Venkatapurum, S., & Coast, J. (2012). The capability approach: an alternative paradigm for health economics? In A. Jones (Ed.), *Elgar companion to health economics*. Cheltenham: Edward Elgar Publishing. (Reprinted from: In File).

Stamuli, E. (2011). Health outcomes in economic evaluation: who should value health? *British Medical Bulletin, 97*, 197-210.

Swait, J. (2006). Advanced choice models. . In B. J. Kanninen (Ed.), *Valuing environmental amenities using stated choice studies. A common sense approach to theory and practice* (pp. 229-293). Dordrecht: Springer Netherlands.

Swait, J., & Bernardino, A. (2000). Distinguishing taste variation from error structure in discrete choice data. *Transportation Research Part B, 34*, 1-15.

Swait, J., & Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research, 30*, 305-314.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34*, 273-286.

Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research, 12*, 599-607.

ACCEPTED MANUSCRIPT

Table 1 – Socio-demographic and end of life characteristics of the sample (n=6020)

Variable	Categories	Frequency	Sample %	Population %*
Gender	Male	2778	46.1	49
	Female	3242	53.9	51
Age	18-24	429	7.1	12
	25-34	901	15.0	8.7
	35-44	1072	17.8	26
	45-54	931	15.5	24.8
	55-64	1177	19.6	7.7
	65+	1510	25.1	20.9
Employment	Employed	2561	42.5	52
	Self-employed	437	7.3	9
	Retired	1793	29.8	14
	Other	1229	20.4	24
Marital Status	Married/Civil partner	3249	54.0	49.4
	Relationship unmarried	986	16.4	8.8
	Single	1480	24.6	34.7
	Widowed	305	5.1	7.0
Religiosity – "*Would you Describe yourself as…*"	Extremely religious	107	1.8	
	Very religious	347	5.8	
	Somewhat religious	1488	24.7	
	Neither religious nor non-religious	1802	29.9	
	Somewhat non-religious	687	11.4	
	Very non-religious	655	10.9	
	Extremely non-religious	869	14.4	
Diagnosed with life limiting illness	Yes	598	9.9	
	No	5422	90.1	
Caring for someone with life limiting illness	Yes	409	6.8	
	No	5611	93.2	
Had someone close die in the last two years	Yes	2459	40.8	
	No	3561	59.2	

* Source: Census 2011 UK Key statistics Tables KS101UK, KS102UK, KS201UK, KS103UK and KS601UK- http://www.nomisweb.co.uk/census/2011/key_statistics_uk , Accessed 7th July 2016.

** The population percentage aged 18-19 years old was imputed assuming that age was equally distributed within the 15-19 age band.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2. Responses to the attributes of ICECAP-A (n = 6020)

ICECAP-A attributes	Frequency	Percentage
Stability		
I am able to feel settled and secure in all areas of my life	1978	32.9
I am able to feel settled and secure in many areas of my life	2801	46.5
I am able to feel settled and secure in a few areas of my life	1003	16.7
I am unable to feel settled and secure in any areas of my life	238	4.0
Attachment		
I can have a lot of love, friendship and support	3162	52.5
I can have quite a lot of love, friendship and support	1989	33.0
I can have a little love, friendship and support	780	13.0
I cannot have any love, friendship and support	89	1.5
Autonomy		
I am able to be completely independent	3789	62.9
I am able to be independent in many things	1730	28.7
I am able to be independent in a few things	423	7.0
I am unable to be at all independent	78	1.3
Achievement		
I can achieve and progress in all aspects of my life	2130	35.4
I can achieve and progress in many aspects of my life	2696	44.8
I can achieve and progress in a few aspects of my life	1037	17.2
I cannot achieve and progress in any aspects of my life	157	2.6
Enjoyment		
I can have a lot of enjoyment and pleasure	2708	45.0
I can have quite a lot of enjoyment and pleasure	2354	39.1
I can have a little enjoyment and pleasure	886	14.7
I cannot have any enjoyment and pleasure	72	1.2

Table 3. Best-worst scale adjusted latent classes with covariates (n=6020)

	Class 1 *Weak preferences: Support & Emotional suffering*		**Class 2** *Dignity, Choice & Support*		**Class 3** *Love, affection, Support & Dignity*		**Class 4** *Support & Physical suffering*	
Support (mean)[1]	0.588	**	0.284	**	0.235	**	0.558	**
Physical Suffering (mean)	0.062		-0.797	**	-0.606	**	-1.733	
Preparation (mean)	0.093	**	0.429	**	0.666	**	0.651	**
Love and Affection (mean)	0.147	**	0.456	**	0.395	**	0.591	**
Choice (mean)	-0.221	**	0.364	**	0.173	**	0.371	**
Emotional suffering (mean)	-0.644	**	-0.616	**	-0.594	**	-0.637	**
Dignity (mean)	-0.025		-0.120	**	-0.269	**	0.200	**
Support_never (1)	-0.529	**	-3.551	**	-4.491	**	-3.713	**
Support_only a little (2)	-0.133	**	-0.544	**	-0.931	**	-0.275	**
Support_ some (3)	0.226	**	1.168	**	1.515	**	1.294	**
Support_most (4)	0.436	**	2.927	**	3.907	**	2.695	**
Physical Suffering_always(1)	-		-1.505	**	-3.076	**	-3.146	**
Physical Suffering_often (2)	-		-0.807	**	-1.783	**	-2.032	**
Physical Suffering_sometimes (3)	-		0.136	**	0.659	**	-0.246	**
Physical Suffering_rarely(4)	-		2.176	**	4.200	**	5.424	**
Preparation_most (1)	-0.246	**	-2.367		-1.994	**	-1.322	**
Preparatio_some (2)	-0.019		-0.137	**	-0.316	**	-0.097	
Preparation_only few (3)	0.112	**	0.651	**	0.526	**	0.330	**
Preparation_not any (4)	0.152	**	1.853	**	1.784	**	1.090	**
Love & Affection_never (1)	-0.360	**	-3.300	**	-5.158	**	-2.668	**
Love & Affection_only a little(2)	-0.120	**	-0.476	**	-1.589	**	-0.321	**
Love & Affection_some(3)	0.203	**	1.047	**	2.203	**	0.915	**
Love & Affection_most (4)	0.277	**	2.729	**	4.544	**	2.074	**
Choice_never(1)	-0.176	**	-3.652	**	-3.272	**	-2.361	**
Choice_only a little(2)	-0.080	**	-1.121	**	-0.996	**	-0.484	**
Choice_some (3)	0.063	*	1.554	**	1.059	**	0.815	**
Choice_most(4)	0.192	**	3.219	**	3.208	**	2.030	**
Emotional suffering_always (1)	-		-1.050	**	-2.642	**	-1.773	**
Emotional suffering_often(2)	-		-0.567	**	-1.771	**	-1.270	**
Emotional suffering_sometimes(3)	-		0.164	**	0.582	**	0.263	**
Emotional suffering_rarely(4)	-		1.452	**	3.831	**	2.780	**
Dignity_never(1)	-0.282	**	-4.889	**	-4.074	**	-3.427	**
Dignity_only a little(2)	-0.053	**	-2.003	**	-1.627	**	-0.920	**
Dignity_some(3)	0.095	**	2.492	**	1.521	**	1.547	**
Dignity_most(4)	0.241	**	4.400	**	4.179	**	2.800	**
Religiosity								
Extremely religious	0.468	**	-0.174		0.145		-0.439	
Very religious	0.012		0.135		-0.040		-0.107	
Somewhat religious	0.067		-0.013		0.007		-0.061	
Neither religious nor non-religious	-0.031		0.098		-0.050		-0.016	
Somewhat non-religious	-0.362	**	0.123		0.217	**	0.022	
Very non-religious	-0.363	**	-0.189	*	0.239	**	0.313	**
Extremely non-religious	-0.102		-0.087		-0.069		0.258	**
Can't choose	0.312		0.107		-0.449		0.031	
Employment								

Employed	0.255	**	-0.202	**	-0.096		0.044	
Self-employed	-0.151		-0.003		0.062		0.093	
Looking after the home and/or family	-0.065		-0.121		0.238	*	-0.052	
Unemployed	0.058		-0.132		0.074		-0.001	
Retired	-0.199	*	-0.031		0.121		0.109	
In full-time education	-0.352	*	0.377		0.005		-0.029	
Long-term sick leave or am disabled	0.214		0.171		0.045		-0.429	**
Full time carer for a friend or relative	0.242		-0.059		-0.448		0.265	
Gender								
Male	0.254	**	-0.195	**	-0.281	**	0.222	**
Female	-0.254	**	0.195	**	0.281	**	-0.222	**
Diagnosed with a life-limiting illness								
Yes	0.130	**	0.024		-0.047		-0.107	*
No	-0.130	**	-0.024		0.047		0.107	*
Caring for someone with life-limiting illness								
Yes	0.093		-0.056		-0.131	*	0.094	
No	-0.093		0.056		0.131	*	-0.094	
ICECAPA*	**-0.318**	****	**0.251**		**0.237**		**-0.170**	
Age*								
16-24	1.176	**	-1.625	**	0.781	**	-0.332	
25-34	0.374	**	-0.495	**	0.321	**	-0.201	*
35-44	-0.131	*	0.043		0.026		0.062	
45-54	-0.270	**	0.246	**	-0.068		0.092	
55-64	-0.665	**	0.837	**	-0.379	**	0.207	**
65+	-0.485	**	0.993	**	-0.680	**	0.172	
Had someone close die in the last two years*								
Yes	0.049		0.075	**	-0.044		-0.079	**
No	-0.049		-0.075	**	0.044		0.079	**
Class membership*	0.642	**	-0.391	*	-0.102		-0.149	
			Coef		s.e.			
Est' scale factor(1)			1		.			
Est' scale factor(2)			0.499	**	0.0058			
Log-likelihood = -255820.7								

*, ** indicates 5% and 1% level of significance.
[1]Attribute means and effects coded indicators;

ACCEPTED MANUSCRIPT

Table 4. Pooled scale adjusted latent classes with covariates (n=6020)

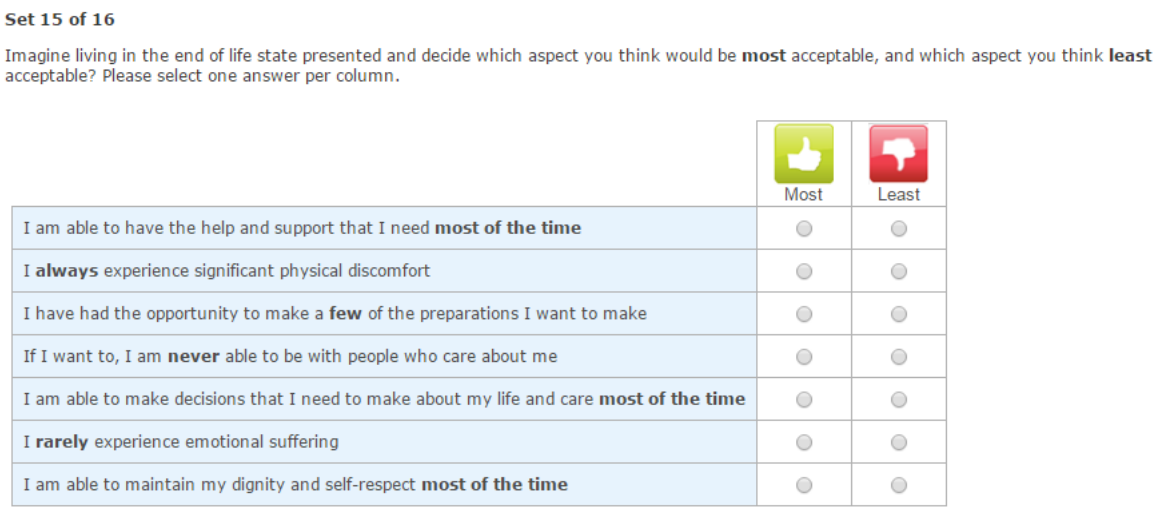
Parameter Estimates	**Class1**		**Class 2**		**Class 3**		**Class 4**	
	Weak preference: Support		*Love, affection & Support*		*Minimising Physical suffering*		*Dignity and Choice*	
ASC (Indicators)								
DCE constant	-0.5718	***	-0.9352	***	-0.5418	***	-0.7693	***
ISupport	0.245	***	0.3965	***	0.2588	***	0.1637	***
IPhysical suffering	-0.2657	***	-0.549	***	0.2013	***	-0.5596	***
IPreparation	-0.0394	***	-0.4223	***	-0.3572	***	-0.2376	***
Ilove&affection	0.1012	***	1.0195	***	0.0088		-0.0243	*
IChoice	-0.0682	***	-0.0197	**	-0.1307	***	0.2317	***
IEmotional suffering	0		0		0		0	
IDignity	0.0749	***	0.1324	***	0.2151	***	0.8591	***
Decrements								
DSupport	0.1669	***	1.0144	***	0.7193	***	0.6819	***
DPhysical Suffering	-0.0167	**	0.6502	***	1.3065	***	0.4201	***
DPreparation	0.08	***	0.462	***	0.2742	***	0.4002	***
Dlove&affection	0.1212	***	1.3341	***	0.5527	***	0.5372	***
DChoice	0.088	***	0.7929	***	0.5112	***	0.7336	***
DEmotional Suffering	0.1889	***	0.9297	***	0.7207	***	0.7016	***
DDignity	0.1381	***	0.9807	***	0.7898	***	1.2356	***
DSupportXDPhysical	0.0176		0.0749	***	0.1421	***	0.1281	***
DSupportXDPreparation	0		0		0		0	
DSupportXDlove	0		0		0		0	
DSupportXDChoice	-0.0013		0.0745	***	0.0874	***	0.0807	***
DSupportXDEmotional	-0.0088		0.0867	***	0.0306	*	0.0114	
DSupportXDDignity	0.0074		0.1479	***	0.0394	***	0.0358	**
DPhysicalXDPreparation	-0.0203	**	-0.0464	***	-0.0087		0.0155	
DPhysicalXDlove	0.0103		0.0343	**	0.0155		-0.0445	***
DPhysicalXDChoice	-0.014		0.1277	***	0.0365	*	0.0427	**
DPhysicalXDEmotional	0.0077		0.0881	***	0.1274	***	0.0524	***
DPhysicalXDDignity	-0.0468	***	-0.0568	***	-0.001		-0.0917	***
DPreparationXDlove	-0.0141		0.0595	***	0.1258	***	0.0658	***
DPreparationXDChoice	-0.0108		-0.0879	***	-0.0723	***	-0.102	***
DPreparationXDEmotional	0.0689	***	0.1502	***	-0.0168		0.082	***
DPreparationXDDignity	0		0		0		0	
DLoveXDChoice	0.0119		0.0508	***	0.0586	***	0.0404	***
DLoveXDEmotional	0		0		0		0	
DLoveXDDignity	-0.0235	**	0.1353	***	0.0713	***	0.073	***
DChoiceXDEmotional	0.0101		0.0835	***	0.1171	***	0.1423	***
DChoiceXDDignity	0		0		0		0	
DEmotionalXDDignity	-0.0017		-0.0053		0.0285	*	0.0476	***
Age 65+ & indicators								
AGE65XISupport	0.0535	***	0.1404	***	0.0328	***	0.0443	***
AGE65XIPhysical	-0.1566	***	-0.1033	***	0.0264	**	-0.0845	***
AGE65XIPrepared	0		0		0		0	
AGE65XILove	0.0272	***	0.0171	*	-0.0517	***	0.0217	**
AGE65XIChoice	0		0		0		0	
AGE65XIEmotional	0		0		0		0	
AGE65XIDignity	0.0825	***	0.1072	***	0.0904	***	0.0943	***
Age 65+ & decrements								
AGE65XDSupport	0.0789	***	0.0482	***	-0.002		0.0463	***
AGE65XDPhysical	-0.0105	**	-0.0616	***	0.0766	***	-0.0239	***
AGE65XDPreparation	0.0577	***	0.0001		-0.0061		0.0228	***
AGE65XDLove	0.0453	***	-0.0518	***	-0.0264	***	0.0281	***
AGE65XDChoice	0.0275	***	-0.0076	*	0.0056		0.0191	***
AGE65XDEmotional	0.0642	***	-0.0078		-0.0289	***	0.0163	***

AGE65XDDignity	0.0781	***	0.0394	***	0.0317	***	0.0719	***
Lifelimiting illness & indicators								
LLXISupport	-0.0132		0.0334	***	-0.0552	***	0.0138	
LLXIPhysical	0.0172	*	-0.1052	***	0.0087		-0.0092	
LLXIPreparation	0		0		0		0	
LLXILove	0.0077		0.1338	***	-0.0189		0.0387	***
LLXIChoice	0		0		0		0	
LLXIEmotional	0		0		0		0	
LLXIDignity	-0.0225	**	-0.0038		0.0398	***	-0.0128	
Lifelimiting illness & decrements								
LLXDSupport	-0.0177	***	0.0092		0.0014		-0.0015	
LLXDPhysical	-0.0159	**	-0.0758	***	0.0035		-0.0235	***
LLXDPreparation	0		0		0		0	
LLXDLove	-0.0026		0.0771	***	0.0263	**	-0.0008	
LLXDChoice	-0.0008		0.0018		0.0269	***	0.0035	
LLXDEmotional	-0.0202	***	0.0198	*	0.0289	***	-0.0096	
LLXDDignity	-0.0146	**	-0.0063		0.0634	***	-0.0257	***
Male & indicators								
MaleXISupport	0.0212	***	0.0681	***	0.0307	***	-0.0362	***
MaleXIPhysical	0.0542	***	0.0543	***	-0.0063		0.0443	***
MaleXIPreparation	0		0		0		0	
MaleXILove	-0.0321	***	-0.1627	***	-0.0568	***	-0.0872	***
MaleXIChoice	0		0		0		0	
MaleXIEmotional	0		0		0		0	
MaleXIDignity	-0.0345	***	0.0232	***	0.012		0.047	***
Male & decrements								
MaleXDSupport	-0.0208	***	0.0225	***	-0.0192	***	-0.0348	***
MaleXDPhysical	-0.0004		0.0536	***	0.0189	***	0.0295	***
MaleXDPreparation	-0.0158	***	-0.0072		-0.0265	***	-0.0092	**
MaleXDLove	-0.0225	***	-0.1242	***	-0.0786	***	-0.0627	***
MaleXDChoice	-0.0108	***	-0.0204	***	-0.0396	***	-0.0128	***
MaleXDEmotional	-0.005	**	-0.0015		-0.027	***	-0.0077	
MaleXDDignity	-0.0252	***	-0.005		-0.0373	***	0.035	***
SCALE parameters								
BWS data	0.7798	***						
BWINDIC	0.0213	***						
Class membership parameters								
ONES	0.0906	**	0.0206		-0.4128	***	0	
GOODNESS-OF-FIT MEASURES								
LL	-301998							
LL0	-412527							
-2*(LLF-LL0)	221058.4	DF=(253) p= 0.000						
#Parameters	253							
Rho-squared	0.268							
Rho-bar-squared	0.267							
AIC	604501.1							
AIC3	604754.1							
BIC	606188.4							
Number of Choice Sets=279312								
Number of Subjects^ =5819								
Tasks per Subject =48								

*, **, *** indicates significance at 10, 5 and 1 % level, respectively. Scale indicators are effects coded.
^ 3.3% respondents removed chose the state on offer more than 13 out of 16 times.

ACCEPTED MANUSCRIPT

Figure 1 – Screenshot of Best-Worst task

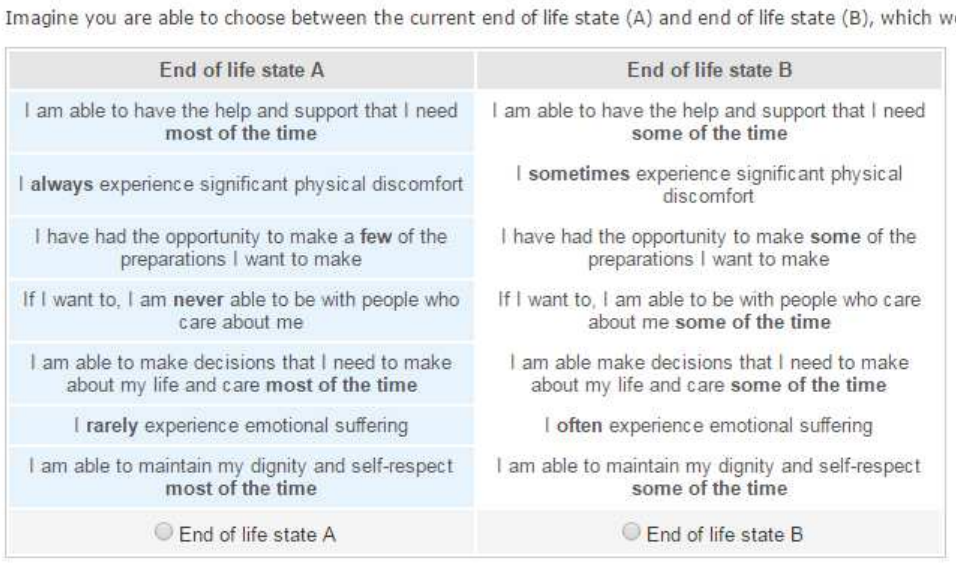


Set 15 of 16

Imagine living in the end of life state presented and decide which aspect you think would be **most** acceptable, and which aspect you think **least** acceptable? Please select one answer per column.

	Most	Least
I am able to have the help and support that I need **most of the time**	○	○
I **always** experience significant physical discomfort	○	○
I have had the opportunity to make a **few** of the preparations I want to make	○	○
If I want to, I am **never** able to be with people who care about me	○	○
I am able to make decisions that I need to make about my life and care **most of the time**	○	○
I **rarely** experience emotional suffering	○	○
I am able to maintain my dignity and self-respect **most of the time**	○	○

Figure 2 – Screenshot of discrete choice task



Imagine you are able to choose between the current end of life state (A) and end of life state (B), which would you choose?

End of life state A	End of life state B
I am able to have the help and support that I need **most of the time**	I am able to have the help and support that I need **some of the time**
I **always** experience significant physical discomfort	I **sometimes** experience significant physical discomfort
I have had the opportunity to make a **few** of the preparations I want to make	I have had the opportunity to make **some** of the preparations I want to make
If I want to, I am **never** able to be with people who care about me	If I want to, I am able to be with people who care about me **some of the time**
I am able to make decisions that I need to make about my life and care **most of the time**	I am able make decisions that I need to make about my life and care **some of the time**
I **rarely** experience emotional suffering	I **often** experience emotional suffering
I am able to maintain my dignity and self-respect **most of the time**	I am able to maintain my dignity and self-respect **some of the time**
○ End of life state A	○ End of life state B

Figure 3(a). Best-worst scale adjusted latent classes – attribute means

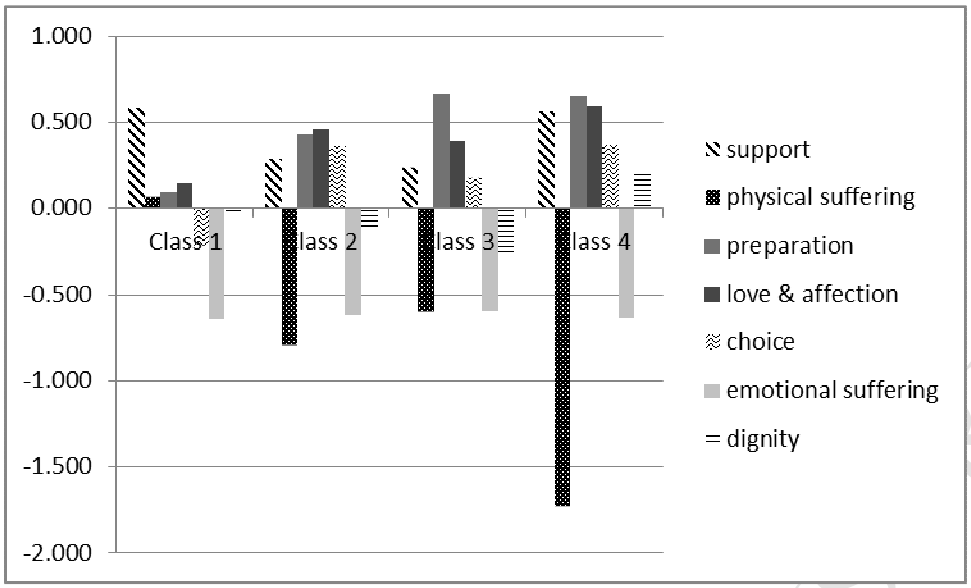


Figure 3(b). Best-worst scale adjusted latent classes – attribute levels

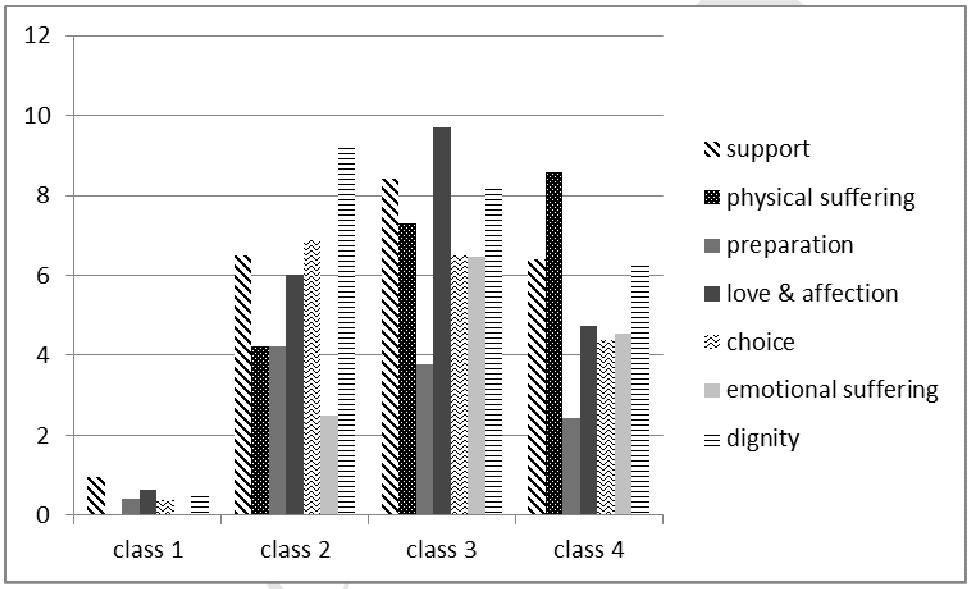
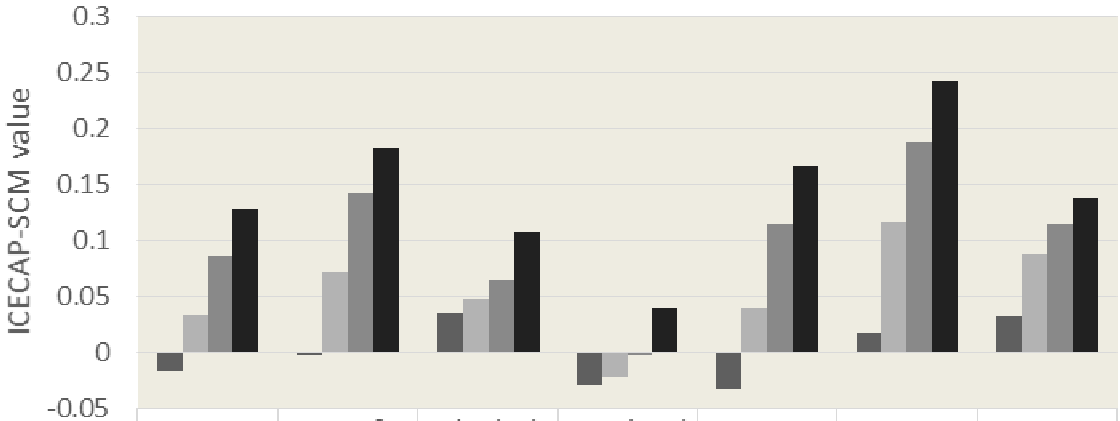


Figure 4. ICECAP-SCM tariff derived from best-worst scaling data



	Choice	Love & Affection	Physical Suffering	Emotional Suffering	Dignity	Support	Prepared
1	-0.017	-0.003	0.034	-0.03	-0.033	0.017	0.032
2	0.033	0.072	0.048	-0.022	0.039	0.116	0.087
3	0.086	0.142	0.064	-0.002	0.115	0.187	0.114
4	0.128	0.181	0.108	0.039	0.166	0.241	0.137

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ACKNOWLEDGEMENTS: We would like to thank all those who participated in the research. We would also like to thank members of the EconEndLife Advisory and Ethics Groups and the Health Economics at Bristol Journal Club, particularly Kathy Armour, Cara Bailey, Alastair Canaway, Louise Jones, Paul Mitchell and Jeff Round, for helpful discussion around the interpretation of the tariffs.

ACCEPTED MANUSCRIPT